REMARKS

The above amendments to the above-captioned application along with the following remarks are being submitted as a full and complete response to the Official Action dated September 30, 2008. In view of the above amendments and the following remarks, the Examiner is respectfully requested to give due reconsideration to this application, to indicate the allowability of the claims, and to pass this case to issue.

Status of the Claims

Claims 1-14, 16 and 24-29 are under consideration in this application. Claims 1, 5, 11-12, 14 and 26-28 are being amended, as set forth in the above marked-up presentation of the claim amendments, in order to more particularly define and distinctly claim Applicants' invention. All the amendments to the claims are supported by the specification. Applicants hereby submit that no new matter is being introduced into the application through the submission of this response.

Prior Art Rejection

Claims 1-5 and 7-23 remain rejected under 35 U.S.C. §102(b) as being anticipated by Chellis et al. (US 6,901,446), and claims 6, 26 and 29 were rejected under 35 U.S.C. §103 (a) as being unpatentable over Chellis '446 in view of Gerszberg et al. (US 6,385,693). These rejections have been carefully considered, but are most respectfully traversed, as more fully discussed below.

A load distribution method of the present invention is adopted by a client-server system 1100 (for example, the embodiment depicted in Fig. 1 or Fig. 11) including a plurality of clients 100/2000, a server cluster 2600 and a storage device 2300 which is coupled to said serves cluster and stores data to be used by at least one of the clients, said server cluster including a plurality of servers 800 each of which includes a cache memory, and said server cluster being used for processing requests related to the data in the storage device from said clients and allowing a number of said servers to be changed dynamically. The method comprises: detecting, by one of clients, a change of the number of servers forming said server cluster (by a server-count detection function 401 of the load control program 400 of Figs. 1-2 in one client 100; [0063] of corresponding US Pub. No. 2005/0038890; Step 1202 in Fig. 5); setting, by said one client 100, an allocation of requests transmissible out to a newly added server 900 at an initial amount smaller than amounts set for the remaining servers in the server cluster, right after detecting an increase in the number of servers (by a connection

distribution function 301 of a load distribution function 300 if the client 100 in Figs. 1 & 9; "*The connection distribution function 301 refers to the connection management table 302 in order to determine a connection is to be allocated to the client program 200 and records the allocation of the connection in the connection management table 302 to reflect the reuse of the connection in the connection management table 302.*" [0084]); monitoring performance of the newly added server; determining whether a cache hit rate for a cache memory in the newly added server is beyond a predetermined value based upon the monitored performance; increasing an amount of requests to be allocated to the newly added server if the cache hit rate is beyond the predetermined value ("*When a server 900 is newly added to the server-cluster system 1100, to which the client 100 transmits out a request, and hence changes the configuration of the server-cluster system 1100 in the client-server system implemented by the first embodiment of the present invention, the number of requests transmissible out to the newly added server 900 is initially set a value small in comparison with that set for each already existing server 800. In this way, it is possible to avoid generation of a long queue of requests each waiting for a processing turn in the newly added server 900 and increase the efficiency of processing in the entire server-cluster system 1100,* ([0088])-[0089])"); transmitting out requests to said servers on the basis of said set allocation, if said increase in the number of servers is detected; and receiving, by said one client, responses to the requests from said servers.

The invention of claim 12 is directed to the client server system of claim 1. The invention of claim 27 is directed one of the client computers of claim 1.

The invention effectively shortens the time that the newly added server 900 takes to process requests, when the new server is added to the server-cluster system ([0024]).

Applicants respectfully submit that none of cited prior art references teaches or suggests such steps of "setting, by said one client, an allocation of requests transmissible out to a newly added server at an initial amount smaller than amounts set for the remaining servers in the server cluster, right after detecting an increase in the number of servers," and then "monitoring performance of the newly added server; determining whether a cache hit rate for a cache memory in the newly added server is beyond a predetermined value based upon the monitored performance; increasing an amount of requests to be allocated to the newly added server if the cache hit rate is beyond the predetermined value" as the present invention.

In contrast, Chellis only allocates resources to be provided to a consumer to improve utilization of the capacity of the one or more resources (Abstract). When a new resource is

discovered, Chellis examines the current allocation of resources, and generates requests to allocate/ reallocate one or more resources based on the availability of the new resource. For example, if a new server with capacity for handle one hundred users comes online, and the current allocation of resources indicates that ten servers each capable of handling fifty users, are currently each handles in excess of forty-five users, Chellis generates reallocation requests to migrate a number of users from one or more of the servers that are operating near capacity to the new server (col. 5, lines 33-46). An amount of resources for an application depends on time (Fig. 4; col. 11, line 59 to col. 12, line 69). Taking an email application as example, as the time passed, the amount of e-mail data is increasing and the email application needs more resources as shown as Figs. 4a-b.

As admitted by the Examiner, Chellis does not disclose a server includes a cache memory and coupled to a storage device (p. 9, last para. of the outstanding Office Action). As such, Chellis does not describe the steps involving a cache hit rate of a cache memory in the newly added server, and balancing the load among the servers accordingly. Gerszberg was relied upon by the Examiner to provide the missing teaching of a cache hit rate of a cache memory.

However, Gerszberg merely locates large caches at the network server platform to reduce traffic on the network trunks or Internet backbone. The NSP cache controller continuously monitors and maintains balanced trance loading among the NSP cache engines. Each NSP cache engine keeps track of the bit rate in each of their cache area responsibilities. Gerszberg only disclose that the NSP cache controller 604 uses access information to dynamically reallocate high traffic addresses from heavily user NSP cache engines to less heavily user NSP cache engines to distribute the traffic load (col. 10, lines 44-52). However, Gerszberg fails to teach that "allocating a smaller amount of requests to the newly added server at first, and increasing the amount of requests to be allocated to the newly added server if the cache hit rate is beyond the predetermined value as determined based upon the monitored performance" as in the present invention.

Regarding the Examiner comments on "allocating a smaller amount of requests to the newly added server at first" (page 3, lines 11-14 of the outstanding Office Action), Applicants respectfully contend that (1) Chellis' adjusting rules by comparing and shifting (col. 5, lines 1-17) simply do not require allocating a smaller amount of requests to the newly added server than the existing servers; and (2) Chellis' description of "a greater or lesser number of resources can be allocated" in col. 11, lines 6-10 as quoted as follows, was

interpreted out of context. The description only said that the numbers of resources/accounts in Fig. 2 are used as examples which can be varied, but not in the context as recite in the claims.

"*While three user accounts 50 are illustrated in connection with FIG. 2, it is to be appreciated that a greater or lesser number of resources (e.g., accounts) can be allocated in accordance with the present invention. Similarly, while ten resources* 25_{A1} *through* 25_{A10} *are illustrated in association with FIG. 2, it is to be appreciated that a greater or lesser number of resources (e.g., processors, disk space) can be allocated in accordance with the present invention.*"

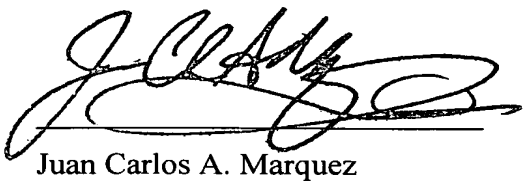
Applicants contend that the cited references and their combinations fail to teach or disclose each and every feature of the present invention as recited in at least independent claims 1, 12 and 27. As such, the present invention as now claimed is distinguishable and thereby allowable over the rejections raised in the Office Action. The withdrawal of the outstanding prior art rejections is in order, and is respectfully solicited.

Conclusion

In view of all the above, clear and distinct differences as discussed exist between the present invention and the prior art references upon which the rejections in the Office Action rely, Applicant respectfully contends that the prior art references cannot anticipate the present invention or render the present invention obvious. Rather, the present invention as a whole is distinguishable, and thereby allowable over the prior art.

Favorable reconsideration of this application is respectfully solicited. Should there be any outstanding issues requiring discussion that would further the prosecution and allowance of the above-captioned application, the Examiner is invited to contact the Applicants' undersigned representative at the address and telephone number indicated below.

Respectfully submitted,



Juan Carlos A. Marquez
Registration Number 34,072

REED SMITH LLP
3110 Fairview Park Drive, Suite 1400
Falls Church, Virginia 22042
(703) 641-4200

February 27, 2009

JCM/JT